Overview of deep learning theory and its application

------- Beijing University of Chemical and Technology's curriculum homework

Author : Xiaoning Liu
Mentor : RuiRui Li
Date:2024.1.6

Abstract

Deep Learning is a new research direction in the field of machine learning, which is introduced into machine learning to make it closer to the original goal -AI(Artificial Intelligence).

Deep learning is the inherent law and level of learning sample data. The information obtained in these learning processes is very helpful for the interpretation of data such as text, images, and sounds. Its ultimate goal is to allow machines to analyze learning ability like humans and can recognize data such as text, images and sounds. It is a complex machine learning algorithm, which has achieved the effect in terms of voice and image recognition, far exceeding the previous related technologies, especially in searching technology, data mining, machine translation, natural language processing, multimedia learning, voice, recommendation and personalized technologies, and other related fields. This article discusses the theoretical knowledge of deep learning and investigates the application of the algorithm in various fields, to provide a certain reference for deep learning studies.

Key Words: Deep learning, natural language processing

Chapter 1 Fundamentals

1.1 Theoretical knowledge

Deep learning is a kind of pattern analysis In terms of specific research contents, methods generally refer to three types of methods:

(1) Neural network system based on convolution operation, namely Convolutional neural network (CNN).

(2) Self coding Based on Multilayer Neurons neural network, including Auto encoder and Sparse coding Sparse Coding.

(3) Pre training is carried out in the form of multi-layer self coding neural network, and then the weight of neural network is further optimized by combining the identification information Deep confidence network(DBN).

Through multi-level processing, the initial "low level" feature representation is gradually transformed into "high level" feature representation, and the "simple model" can be used to complete complex classification and other learning tasks. Thus, deep learning can be understood as "feature learning" or "representation learning".

In the past, when machine learning was used for real tasks, the features used to describe samples usually needed to be designed by human experts, which became "feature engineering". As we all know, the quality of features has a crucial impact on generalization performance, and it is not easy for human experts to design good features. In feature

learning, Good features are generated by machine learning technology itself, which makes machine learning move forward to "fully automatic data analysis".

In recent years, researchers have gradually combined these methods to improve the prediction accuracy. Compared with the traditional learning method, deep learning method presupposes more model parameters, so the model training is more difficult. According to the general rule of statistical learning, the more model parameters, the more data to participate in training.

In the 1980s and 1990s, due to the limited computing power of computers and the limitations of related technologies, the amount of data available for analysis was too small, and deep learning did not show excellent recognition performance in pattern analysis. Since 2006, Hinton and others have proposed that fast computing is limited by Boltzmann machine(RBM) and biased CD-K algorithm, RBM has become a powerful tool to increase the depth of neural networks, leading to the emergence of deep networks such as widely used DBN (developed by Hinton and used by Microsoft and other companies in speech recognition). At the same time, sparse coding is also used in deep learning because it can automatically extract features from data. The convolutional neural network method based on local data region has also been extensively studied in recent years.

1.2 Content

Deep learning is a type of machine learning, and machine learning is a necessary path to achieve artificial intelligence. The concept of deep learning originated from the research of artificial neural networks, and a multi-layer perceptron with multiple hidden layers is a type of deep learning structure. Deep learning combines low-level features to form more abstract high-level representations of attribute categories or features, in order to discover distributed feature representations of data. The motivation for studying deep learning lies in establishing neural networks that simulate the human brain for analytical learning, which mimic the mechanisms of the human brain to interpret data such as images, sounds, and texts.

The computation involved in generating an output from an input can be represented by a flow graph: a flow graph is a graph that represents computation, in which each node represents a basic computation and a computed value, and the computed result is applied to the values of the child nodes of that node. Consider a computational set that can be allowed at each node and possible graph structure, and define a family of functions. The input node has no parent node, and the output node has no child nodes. A special property of this flow chart is depth: the length of the longest path from one input to one output.

Traditional feedforward neural networks can be seen as having a

depth equal to the number of layers.

One of the directions of artificial intelligence research is represented by the so-called "expert system", which is defined by a large number of "If Then" rules and follows a top-down approach. Artificial Neural Network marks another bottom-up approach. Neural networks do not have a strict formal definition. Its basic feature is to attempt to mimic the patterns of information transmission and processing between neurons in the brain.

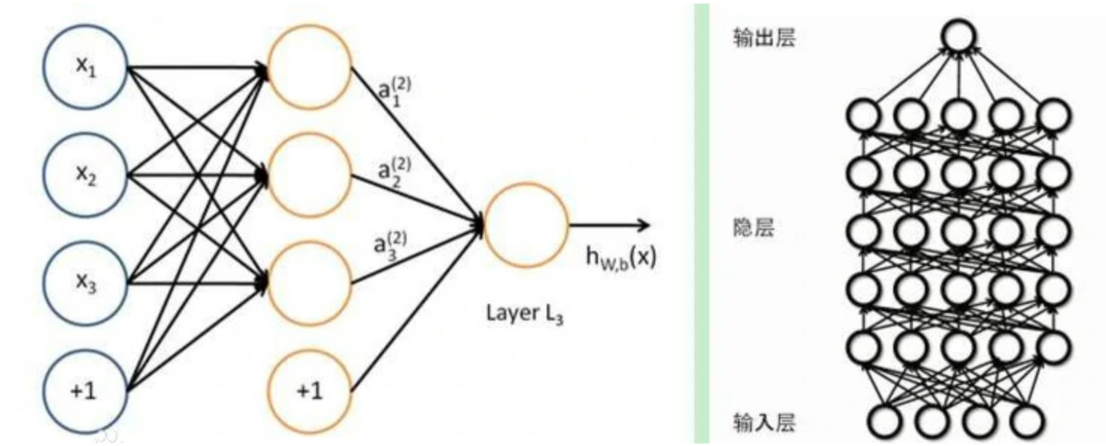

Figure1 A Deep Learning Model with Multiple Hidden Layers

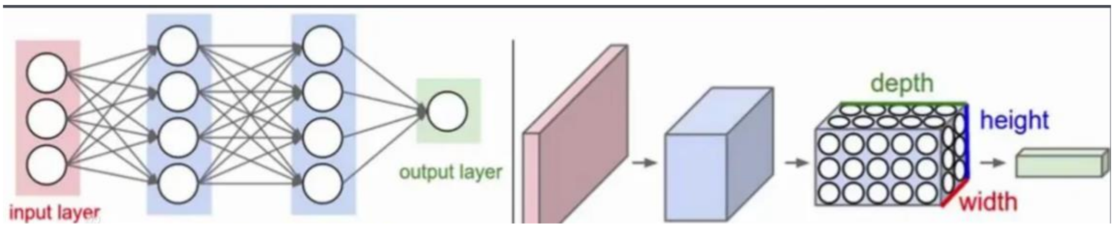

Figure2 The training process

1.3 Characteristic

Compared to traditional shallow learning, deep learning has the follow characteristics. Firstly, in deep learning, Emphasis is placed on the depth of the model structure, typically consisting of 5, 6, or even 10

layers of hidden nodes. Secondly, The importance of feature learning has been clarified. That is to say, through layer-by-layer feature transformation, the feature representation of the sample in the original space is transformed into a new feature space, making classification or prediction easier. Compared with the method of manually constructing features based on rules, using big data to learn features can better characterize the rich intrinsic information of the data.

By designing and establishing an appropriate number of neural computing nodes and multi-layer operation hierarchies, selecting appropriate input and output layers, and through network learning and optimization, a functional relationship from input to output is established. Although the functional relationship between input and output cannot be found absolutely, it can approximate the real correlation relationship as much as possible. By using a successfully trained network model, we can achieve our automation requirements for complex transaction processing.

Chapter 2 Typical Models

2.1 Convolutional neural network model

Before the emergence of unsupervised pre-training, training deep neural networks was often very difficult, and one special case was convolutional neural networks. Convolutional neural networks are inspired by the structure of the visual system. The first convolutional neural network computational model was proposed in Fukushima's neurocognitive machine. Based on local connections between neurons and hierarchical image transformation, neurons with the same parameters are applied to different positions in the previous layer of the neural network to obtain a translation invariant neural network structure. Later, Le Cun et al. designed and trained convolutional neural networks using error gradients based on this idea, achieving superior performance in some pattern recognition tasks. So far, the pattern recognition system based on convolutional neural networks is one of the best implementation systems, especially in handwritten character recognition tasks, showing extraordinary performance.

The connections between convolutional layers in convolutional neural networks are called sparse connections, meaning that compared to fully connected feedforward neural networks, the neurons in the convolutional layers are only connected to a portion of their adjacent layers, not all of them. Specifically, any pixel (neuron) in the l-th layer

feature map of a convolutional neural network is only a linear combination of pixels within the receptive field defined by the convolutional kernel in the l-1 layer. The sparse connections of convolutional neural networks have the effect of regularization, improving the stability and generalization ability of the network structure, avoiding overfitting. At the same time, sparse connections reduce the total number of weight parameters, which is beneficial for the fast learning of neural networks and reduces memory consumption during computation.

In convolutional neural networks, all pixels in the same channel of the feature map share a set of convolutional kernel weight coefficients, a property known as weight sharing. Weight sharing distinguishes convolutional neural networks from other neural networks that contain local connection structures. Although the latter uses sparse connections, the weights of different connections are different. Weight sharing, like sparse connections, reduces the total number of parameters in convolutional neural networks and has a regularization effect.

From the perspective of fully connected networks, the sparse connections and weight sharing of convolutional neural networks can be regarded as two infinitely strong priors, that is, all weight coefficients of a hidden layer neuron outside its receptive field are always zero, but the receptive field can move in space. And within one channel, all neurons have the same weight coefficients.

2.2 DBN Model

DBN can be explained as a Bayesian probability generation model, consisting of multiple layers of random latent variables. The upper two layers have undirected symmetric connections, while the lower layer obtains a top-down directed connection from the previous layer. The state of the lowest level unit is a visible input data vector. DBN consists of a stack of 2F structural units, typically RBM. The number of visible layer neurons in each RBM unit in the stack is equal to the number of hidden layer neurons in the previous RBM unit. According to the deep learning mechanism, the input examples are used to train the first layer of RBM units, and the output is used to train the second layer of RBM models. The RBM models are stacked by adding layers to improve model performance. In the unsupervised pre training process, DBN encodes the input to the top-level RBM, decodes the top-level state to the lowest level unit, and achieves input reconstruction. RBM, as the structural unit of DBN, shares parameters with each layer of DBN.

2.3 Stacked Self Encoding Network Model

The structure of a stack based self-coding network is similar to that of a DBN, consisting of several stacked structural units. The difference is that its structural units are self-coding models (auto encoder) rather than RBM. The self-coding model is a two-layer neural network, with the first layer called the encoding layer and the second layer called the decoding

layer.

Chapter 3 Application

Deep learning has extensive applications in various fields such as computer vision, speech recognition, and natural language processing. The Multimedia Laboratory of The Chinese University of Hong Kong was one of the earliest Chinese teams to apply deep learning in computer vision research. At the world-class artificial intelligence competition LFW (Large scale Facial Recognition Competition), the laboratory once surpassed FaceBook to win the championship, making the recognition ability of artificial intelligence in this field surpass that of real people for the first time.

Microsoft researchers first introduced RBM and DBN into speech recognition acoustic model training through collaboration with Hinton, and achieved great success in large vocabulary speech recognition systems, resulting in a relative reduction of 30 percent in speech recognition error rate. However, DNN does not yet have an effective parallel fast algorithm, and many research institutions are using large-scale data corpus to improve the training efficiency of DNN acoustic models through GPU platforms. Internationally, companies such as IBM and Google have rapidly conducted research on DNN speech recognition, and the speed is extremely fast. In China, companies or research units such as Alibaba, Baidu, and the Institute of Automation of the Chinese Academy of Sciences are also conducting research on deep learning in

speech recognition.

Many institutions have conducted research in the field of natural language processing. In 2013, Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean published a paper titled Efficient Estimation of Word Representations in Vector Space to establish a word2vector model. Compared with traditional bag of words models, word2vector can better express grammatical information. Deep learning is mainly applied in fields such as natural language processing, machine translation, and semantic mining.

In 2020, deep learning can accelerate innovation in semiconductor packaging and testing. In terms of reducing repetitive labor, improving yield, controlling accuracy and efficiency, and reducing detection costs, AI deep learning driven AOI has broad market prospects, but it is not easy to handle. On April 13, 2020, in a medical and artificial intelligence (AI) study published in the British journal Nature Machine Intelligence, Swiss scientists introduced an AI system that can scan cardiovascular blood flow within seconds. This deep learning model is expected to enable clinical physicians to observe real-time changes in blood flow while patients undergo MRI scans, thereby optimizing the diagnostic workflow.

References

[1] Max C, Christian V,Tobias S, et al. Deep Concrete Flow: Deep learning based characterisation of fresh concrete properties from open-channel flow using spatio-temporal flow fields[J].Construction and Building Materials,2024,411134809-.

[2] Xiaofei Y, Enrique C R D ,Yang Z, et al. UAV-deployed deep learning network for real-time multi-class damage detection using model quantization techniques[J].Automation in Construction,2024,159105254-.

[3] Jiyu X, Mitsuyoshi A, M.DF. Autonomous detection of steel corrosion spatial variability in reinforced concrete using X-ray technology and deep learning-based semantic segmentation[J]. Automation in Construction, 2024,158105252-.

[4] Sheng J, Xinming W, Qinghao M. Spatial memory-augmented visual navigation based on hierarchical deep reinforcement learning in unknown environments[J]. Knowledge-Based Systems, 2024, 285111358-.

[5] Xukai Z, Yuxing L, Guangsi L. An integrated deep learning approach for assessing the visual qualities of built environments utilizing street view images[J]. Engineering Applications of Artificial Intelligence, 2024, 130107805-.

[6] Zhaohui L, Longyan W, Jian X, et al. A deep learning framework for reconstructing experimental missing flow field of hydrofoil[J]. Ocean Engineering, 2024, 293116605-.

[7] Ruoning Y, Wenyi R, Man Z , et al. Transparent objects segmentation based on polarization imaging and deep learning[J]. Optics Communications, 2024, 555130246-.
[8] Mina K, Angelo F, Ludovic L. Improving deep-learning methods for area-based traffic demand prediction via hierarchical reconciliation[J]. Transportation Research Part C, 2024, 159104410-.